# An optimal investor state dispute settlement mechanism*

Frank Stähler†

Version of April 21, 2016.

**Abstract**

Investor state dispute settlements (ISDS) are supposed to protect a foreign investor against opportunistic behavior of a host country. This paper scrutinizes the optimal design of ISDS provisions that solve the holdup problem. It shows that an efficient investor protection mechanism requires a multilateral framework provided by a supranational institution. Furthermore, any ISDS compensation from the government to the investor must not be based on reductions in investor profits but on the host country's benefits.

**Keywords:** Investor State Dispute Settlement; Foreign Direct Investment; TTIP; TPP.

**JEL Classification:** F21; F23; F53; F55; D82.

*Part of this paper was written while I was a Fred Gruen Visiting Professor at the Australian National University in 2015. I wish to thank the ANU for its hospitality.

†University of Tübingen, University of Adelaide and CESifo, Mohlstr. 36 (V4), D-72074 Tübingen, Germany. Email: frank.staehler@uni-tuebingen.de

# 1 Introduction

Investor state dispute settlements (ISDS) provisions seem to be the most controversial policy issue in international trade policies in recent times. ISDS provisions are blamed to undermine national sovereignty as they allow foreign investors to take any kind of apparently unfair treatment to a tribunal. This tribunal may rule that the host country of the investment will have to indemnify the foreign investor if they find that the host country government policies have caused "unjustified" harm equivalent to expropriation. While ISDS provisions can be found in many bilateral investment treaties (see OECD, 2012), the political opposition against ISDS provisions has gained momentum when it became clear that they are supposed to become an integral part of both the Transatlantic Trade and Investment Partnership (TTIP) presently negotiated between the US and the EU and the Trans-Pacific Partnership (TPP) Agreement presently to be ratified by several Pacific rim countries, including the US, Canada, Australia and Japan. For example, Article 9.8 of the TPP draft (2016) specifies that "[n]o Party shall expropriate or nationalize a covered investment either directly or indirectly through measures equivalent to expropriation or nationalization (expropriation) ..." A further concern is that national governments, anticipating potential compensations, will constrain themselves and may not pursue any policy that might affect future multinational profits, leading to lower regulatory standards in host countries.

The economic reasoning behind ISDS provisions is the well-known holdup problem that may arise in the context of foreign direct investment (FDI). Due to incomplete contracts, a government cannot credibly promise investor-friendly policies if the investment is very specific such that it has little or no value when relocated to another country.[1] As a consequence, some beneficial investment may not take place as the foreign investor will correctly anticipate her exposure to the holdup problem. Interestingly, though, it is not clear whether the recent design of ISDS provisions will lead to more FDI activities.

[1] See Navaretti and Venables (2006) for an exposition of the standard holdup problem. The allocation of property rights can mitigate the holdup problem as shown by Antràs (2003), Antràs and Helpman (2004) and Antràs and Chor (2013).

While Egger and Merlo (2012) can show that bilateral investment treaties increase investment, the impact of ISDS provisions on FDI activities is unclear (see Berger *et al*, 2011). Furthermore, Kohler and Stähler (2016) find that an ISDS provision may improve aggregate welfare generated by a relationship between a foreign investor and a host country if strategic ownership changes could be ruled out. However, the ISDS provision will lead to further distortions and can never achieve the first best. They conclude that a national treatment provision may lead to a better outcome in terms of investor protection. Aisbett *et al* (2000a) show that taking out some well-defined policies from any potential compensation claim (so-called"police powers carve-outs") can improve even the host country's welfare.

Overall, both the empirical and the sparse theoretical literature have discussed the issue of investor protection only by looking upon the implications of past provisions and the expected effects of ISDS provisions for future multilateral agreements. This paper will not take ISDS provisions as given, but will develop two different types of an optimal mechanism that is able to solve the holdup problem. Consequently, the innovation of this paper is that it outlines how an ISDS provision should be designed that may not only mitigate the holdup problem, but will solve it. The paper will show that such a mechanism must involve at least three parties, the investor, the government and an arbitrator. Since the mechanism does not work without arbitrator, the paper emphasizes that a proper investor protection mechanism must be able to rely on a supranational institution. Therefore, the paper will also make a contribution on the role of multilateralism versus bilateralism for investment liberalization.[2] The only paper that considers the efficient compensation of (domestic and foreign) investors is Aisbett *et al* (2010b), but their model assumes a court that receives a stochastic signal.[3] The arbitrator in our model does not receive any signal, but manages an arbitration process according to rules the investor and the government

[2]See for the role of multilateralism versus bilateralism in the context of trade liberalization, Bagwell and Staiger (2005) and Bagwell *et al* (2016).

[3]Furthermore, the regulator in this model has only the choice to shut down the multinational operation. There is also a related literature on compensations for land takings, see for example, Blume *et al* (1984), Hermelin (1995) and Nosal (2001).

have agreed upon.

The remainder of the paper is organized as follows. Section 2 will set up a simple model that will imply the holdup problem in the absence of any investor protection. Section 3 will develop two different designs of an optimal ISDS mechanism, and Section 4 will present an example. Section 5 will offer some concluding remarks.

# 2 The model

The model is an extension of the standard models of foreign direct investment that is specific and potentially subject to opportunistic behavior. The innovation is that the future environment as it affects domestic welfare is *ex ante* uncertain, and that we consider three agents in this model: the domestic government, the foreign investor, and an arbitrator. Prior to the game played and scrutinized here, there has been an investment stage that we do not consider in detail. Instead, we assume for now that the investor has entered as a foreign multinational, but it should be clear that the outcome of the game we consider will determine the overall profitability of the investment. In particular, we assume that the government can exercise an activity after entry, denoted by $a, a \in [0, \infty[$, that will harm the investor and benefit the domestic country.[4] A higher activity level is equivalent to a higher level of intervention or an increase in regulation affecting the investor.

In detail, the investor is assumed to maximize her profits, and the maximized investor profit, denoted by $\pi$, depends negatively on the activity level $a$:

$$\pi(a) : \pi_a(a) < 0, \pi_{aa}(a) \leq 0, \exists \bar{a} > 0 : \pi(\bar{a}) = \bar{\pi} \geq 0. \tag{1}$$

The profit level $\bar{\pi}$ gives the value of the investment when leaving the country. According to (1), an activity level $\bar{a}$ exists such that investor profit will fall below $\bar{\pi}$ for all $a > \bar{a}$. If the investment is completely country-specific such that it is of no value at any other

[4]For the sake of simplicity, we assume $a$ to be a differentiable scalar, but our results do also hold if it were a vector of activities.

location, $\bar{\pi} = 0$, and $a > \bar{a}$ would make investor profit become negative. In any case, the investor will leave the domestic country if $a > \bar{a}$.

Domestic welfare depends on the activity level itself and on the realization of a stochastic variable $\theta$ that measures the degree of intervention necessity. Importantly, while this realization becomes known at least to the government and the investor, it cannot be verified by the arbitrator or any other third party. Thus, the investor and the government cannot agree on any contract or agreement that has to rely on the realization of this stochastic variable, and without any arbitration managed by a supranational institution, the potential relationship will be dominated by a serious holdup problem. In particular, domestic welfare, denoted by $W$, is given by

$$W = \begin{cases} 0 & \text{if } \pi(a) < \bar{\pi}, \\ v(\theta, a) > 0 & \text{if } \pi(a) \geq \bar{\pi}, \end{cases} \tag{2}$$

where

$$v(\theta, a) : v_a(\theta, 0) > 0, v_{aa}(\cdot) < 0, v_{\theta a}(\cdot) > 0, \exists \tilde{a}(\theta) > 0 : v_a(\theta, \tilde{a}(\theta)) = 0.$$

Note that we do not need any assumption on $v_\theta(\cdot)$ for the subsequent analysis. A natural assumption would be that $v_\theta(\cdot)$ is negative such that a large realization of $\theta$ leads to a large direct drop in domestic welfare. What is more important, however, is that marginal welfare improves with $\theta$ such that the domestic government will want to fight a large $\theta$-shock with a large $a$. As for $\theta$, both the government and the investor anticipate that $\theta \in [\underline{\theta}, \overline{\theta}], 0 < \underline{\theta} < \overline{\theta} < \infty$, and they know that $\theta$ is distributed according to the c.d.f. $G(\theta)$. Expression (2) acknowledges that domestic welfare can be strictly positive only if the investor decides to stay in the country. Furthermore, a function $\tilde{a}(\theta)$ exists that maximizes domestic welfare for any realization of $\theta$ for which

$$\tilde{a}'(\theta) = -\frac{v_{\theta a}(\theta, \tilde{a}(\theta))}{v_{aa}(\theta, \tilde{a}(\theta))} > 0,$$

holds, that is, domestic activities increase with the intervention necessity.

What happens without any investor protection? Without any investor protection, the government and the investor play a two-stage game: in the first stage, the government sets $a$, and in the second stage, the investor decides whether to leave or to stay. Since the domestic government has no interest to make the investor leave the country, it will maximize $v(\theta, a)$ with respect to the activity level $a$ subject to $\pi(a) \geq \bar{\pi}$. If $\bar{a} < \tilde{a}(\theta)$, the government will set $a = \bar{a}$, and it may happen that the investor, anticipating this outcome, will not make any investment in a preceding investment stage. In particular, if $\bar{\pi} = 0$, that is, if the investment is completely country-specific, the government *de facto* expropriates the foreign investor. If $\bar{a} \geq \tilde{a}(\theta)$, the government will set $a = \tilde{a}(\theta)$, so the investor will still realize a profit, but this profit may not be sufficient to support entry in the investment stage. Therefore, without any investor protection, the model reproduces the classic holdup problem.

What would the investor and the government do if a third party could verify the intervention necessity $\theta$ and enforce a contract between the two parties? Let aggregate welfare, that is, the sum of domestic welfare and foreign profit, be denoted by $\Omega = W + \pi$. We assume that condition

$$v_a(\underline{\theta}, 0) + \pi_a(0) > 0 \tag{3}$$

holds. Condition (3) guarantees that some $a > 0$ is socially desirable. If both parties could agree on an enforceable investor protection contract, they would specify an action plan $a^*(\theta)$ such that aggregate welfare is maximized, that is, that

$$\Omega_a(\theta, a^*(\theta)) = v_a(\theta, a^*(\theta)) + \pi_a(a^*(\theta)) = 0$$

holds for all $\theta \in [\underline{\theta}, \overline{\theta}]$. Note that $a^*(\theta)$ is a function given the concavity of both the domestic welfare function and the profit function, and that $a^*(\theta) < \tilde{a}(\theta)$. Furthermore, we assume that $a^*(\overline{\theta}) < \bar{a}$ so that the optimal policy will never want the foreign investor to leave the country. We also find that

$$a^{*\prime}(\theta) = -\frac{v_{\theta a}(\theta, a^*(\theta))}{v_{aa}(\theta, a^*(\theta)) + \pi_{aa}(a^*(\theta))} > 0,$$

holds, that is, that the optimal policy will lead to an increase in the domestic activity level with the intervention necessity, as to be expected. If the optimal policy could be implemented, the foreign investor would correctly anticipate that her expected profit at this stage would be equal to $\int_{\underline{\theta}}^{\overline{\theta}} \pi(a^*(\theta))dG(\theta)$, and the entry decision would also be optimal at the investment stage. We now investigate how this optimal policy can be implemented using arbitration by a supranational organization that cannot verify the intervention necessity.

# 3 Designing an optimal ISDS mechanism

The proposed mechanism is an extension and modification of a mechanism suggested by the implementation literature (see Moore and Repullo, 1988, and Maskin and Tirole, 1989).[5] The basic idea is that the government will announce an intervention necessity, denoted by $\hat{\theta}$, and that this announcement cannot be proven true or false by any third party. However, the investor, being familiar with the implications of her investment, can challenge this announcement, and any challenge will have implications for both the investor and the government that will be managed by the arbitrator. In particular, suppose that the government and the investor agree upon the following mechanism to be managed by the arbitrator:

**ISDS mechanism I**

- Stage 1: The government and the investor agree on the optimal action plan $a^*(\hat{\theta})$, on alternative action plans $a_2(\hat{\theta})$ and $a_1(\hat{\theta})$ and transfer schemes $F_1, F_2, T(\hat{\theta})$. The alternative action plans fulfil the following requirements. $\forall \hat{\theta} \in [\underline{\theta}, \overline{\theta}]$ :

[5]The seminal model has been developed in the context of incomplete contracts between private agents like firms and not between a government and a firm. Furthermore, agents in our model do not take any actions right at the start as in Maskin and Tirole (1989), but both only observe the realization of the intervention necessity.

(i) $a^*(\hat{\theta}) \geq a_2(\hat{\theta}) > a_1(\hat{\theta})$, and

(ii) $a_2'(\hat{\theta}), a_1'(\hat{\theta}) > 0$.

- Stage 2: The government announces $\hat{\theta}$.[6]
- Stage 3: The investor may challenge the announcement $\hat{\theta}$.
  - If she does not, action plan $a^*(\hat{\theta})$ is implemented, no transfers are paid and the game is over.
  - If she does, the government has to make an upfront payment of size $F_1$ to the arbitrator.
- Stage 4: If $\hat{\theta}$ has been challenged, the government is offered to continue with action plan $a_1(\hat{\theta})$ for which it will receive a transfer $T(\hat{\theta})$ from the arbitrator. Otherwise action plan $a_2(\hat{\theta})$ is implemented.
  - If the government accepts, action plan $a_1(\hat{\theta})$ is implemented, the government receives $T(\hat{\theta})$ from the arbitrator and the investor receives $F_1 - T(\hat{\theta})$ from the arbitrator.
  - If the government rejects, $a_2(\hat{\theta})$ is implemented and the investor has to pay $F_2$ to the arbitrator.

Some comments are in order now. First, our restrictions on the choice of the alternative action plans are very mild. Second, note carefully that the arbitrator has at no point in this process to guess what the true realization of $\theta$ is. Third, if the government decides to continue with action plan $a_1(\hat{\theta})$ after being challenged, the investor will receive a compensation, similar to existing ISDS provisions. However, we will now show that this compensation must not be based on foregone profits.

We now prepare our first main result. Suppose that the transfer in Stage 4 to the government when accepting $a_1(\hat{\theta})$ is given by

[6] Alternatively, the government can also announce the activity level $a^*$. Since $a^{*\prime}(\hat{\theta}) > 0$, announcing an activity level is equivalent to announcing an intervention necessity.

$$T(\hat{\theta}) = v(\hat{\theta}, a_2(\hat{\theta})) - v(\hat{\theta}, a_1(\hat{\theta})) > 0. \tag{4}$$

$T(\hat{\theta})$ depends only on the announcement, and since $v_a(\cdot) > 0$ and $a_2(\hat{\theta}) > a_1(\hat{\theta})$, $T(\hat{\theta})$ is unambiguously positive. Let $\bar{T} = \max_{\hat{\theta}\in[\underline{\theta},\overline{\theta}]} T(\hat{\theta}) > 0$ denote the maximum transfer. Furthermore, we introduce

$$\begin{aligned}
\Psi^I &= \max_{\hat{\theta}\in[\underline{\theta},\overline{\theta}]} \left\{\pi(a_2(\hat{\theta})) - \pi(a^*(\hat{\theta}))\right\} \geq 0 \text{ and} \\
\Psi^R &= \max_{\theta,\hat{\theta}\in[\underline{\theta},\overline{\theta}]} \left\{v(\theta, a_1(\hat{\theta})) - v(\theta, a^*(\theta))\right\}.
\end{aligned}$$

The terms here reflect the conditions an optimal ISDS mechanism has to meet: the superscript $I$ will refer to the condition that an inappropriate challenge should not to be made, and the superscript $R$ refers to the condition that the government should reveal and announce the true intervention necessity such that $\hat{\theta} = \theta$. Note that $\Psi^I$ is positive as it is the difference in profits between the disagreement plan $a_2(\hat{\theta})$ and the optimal action plan for which $a_2(\hat{\theta}) \leq a^*(\hat{\theta})$.[7] $\Psi^R$ measures the difference between domestic welfare of the action plan after an over-reported intervention necessity has been successfully challenged and domestic welfare of truthfully announcing the intervention necessity. The sign of $\Psi^R$ depends on the domestic welfare function and the design of the alternative action plan $a_1(\hat{\theta})$. It is likely to be positive, but we cannot rule out that $\Psi^R < 0$ if the difference in activity levels between the alternative action plan $a_1(\hat{\theta})$ and the optimal action plan $a^*(\hat{\theta})$ is very large.[8] We find:

**Proposition 1.** *If*

[7]However, it is not clear for which announcement $\hat{\theta}$ is will be largest, as differentiation of $\pi(a_2(\hat{\theta})) - \pi(a^*(\hat{\theta}))$ with respect to $\hat{\theta}$ leads to $\pi'(a_2(\hat{\theta}))a_2'(\hat{\theta}) - \pi'(a^*(\hat{\theta}))a^{*\prime}(\hat{\theta})$ and shows that the change depends on the relative change of the optimal action plan compared to the disagreement action plan if $a_2(\hat{\theta}) < a^*(\hat{\theta})$.

[8]If $v_a(\theta, a_1(\overline{\theta})) \geq 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, $v(\theta, a_1(\hat{\theta}))$ will be largest for $\hat{\theta} = \overline{\theta}$. In this case, $\Psi^R < 0$ if $v(\theta, a_1(\overline{\theta})) < v(\theta, a^*(\theta))$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. This may happen if $a_1(\overline{\theta}) < a^*(\underline{\theta})$ such that the alternative action plan $a_1$ specifies a lower activity level for the largest possible $\hat{\theta}$-announcement than the optimal activity plan $a^*$ does for the smallest $\theta$-realization, that is, for $\theta = \underline{\theta}$.

- $T(\hat{\theta})$ *is set according to* (4),
- $F_2 \geq \Psi^I$ *, and*
- $F_1 \geq \bar{T} + \Psi^R$ *if* $\Psi^R \geq 0$ *or* $F_1 \geq \bar{T}$ *if* $\Psi^R < 0$,

*ISDS mechanism I will imply the optimal action plan* $a^*(\theta)$ *as a subgame-perfect equilibrium.*

We will now develop the proof of Proposition 1. In order to qualify for an optimal mechanism, the mechanism has to meet several conditions. Since $v_a(\theta, a^*(\theta)) > 0$ and $a^{*\prime}(\cdot) > 0$, it is obvious that the government has no incentive to under-report $\theta$: if $\hat{\theta} < \theta$, aggregate welfare would be lower than maximal, and it would only be the investor benefiting from this under-reporting while domestic welfare would decline. Hence, an announcement should be challenged only if the government over-reports $\theta$. Following Aghion *et al* (2015), we label these conditions appropriate challenge and inappropriate challenge, that is, a challenge by the investor is appropriate if $\hat{\theta} > \theta$ and is inappropriate otherwise. Suppose that the investor has challenged the announcement of the government. The transfer scheme $T(\hat{\theta})$ is designed such that only appropriate challenges will be accepted so that $a_1(\hat{\theta}))$ will be implemented only after an appropriate challenge. In particular, the government will accept the offer to continue with action plan $a_1(\hat{\theta})$ if only if

$$\begin{aligned} v(\theta, a_1(\hat{\theta})) + T(\hat{\theta}) &> v(\theta, a_2(\hat{\theta})) \Leftrightarrow \\ v(\hat{\theta}, a_2(\hat{\theta})) - v(\hat{\theta}, a_1(\hat{\theta})) &> v(\theta, a_2(\hat{\theta})) - v(\theta, a_1(\hat{\theta})) \Leftrightarrow \\ \hat{\theta} &> \theta, \end{aligned} \tag{5}$$

where the last line follows from $v_{\theta a}(\cdot) > 0$.[9] Consequently, the offer is accepted if and only if the government has been over-reporting $\theta$. Thus, an inappropriate challenge will never imply the alternative action plan $a_1(\hat{\theta})$.

[9]In detail, let $\phi(x) = v(x, a_2(\hat{\theta})) - v(x, a_1(\hat{\theta}))$. Then, $\phi'(x) = v_x(x, a_2(\hat{\theta})) - v_x(x, a_1(\hat{\theta})) > 0$ because $v_{xa}(\cdot) > 0$ and $a_2(\hat{\theta}) > a_1(\hat{\theta})$. Thus, $\phi(x)$ increases with $x$ and $\phi(x) > \phi(\theta)$ for all $x > \theta$.

We now turn to the investor's incentives to challenge the government's announcement. She should do so in case of over-reporting and will correctly anticipate that the offer to continue with action plan $a_1(\hat{\theta})$ will be accepted. An appropriate challenge will be made if

$$\pi(a_1(\hat{\theta})) + F_1 - T(\hat{\theta}) \geq \pi(a^*(\hat{\theta})). \tag{6}$$

Condition (6) is always fulfilled because $F_1 \geq \bar{T} \geq T(\hat{\theta})$ and $\pi(a_1(\hat{\theta})) > \pi(a^*(\hat{\theta}))$, and consequently, the investor will always challenge the government when it is appropriate.[10] However, this still leaves us with the possibility that the investor will also challenge the government when it is not appropriate. She will not do so if

$$\pi(a_2(\hat{\theta})) - F_2 \leq \pi(a^*(\hat{\theta})), \tag{7}$$

and this condition is fulfilled for all possible announcements due to the specification of $F_2$ that makes the disagreement action plan $a_2(\hat{\theta})$ unattractive for the investor. Note that $a_2(\hat{\theta}) = a^*(\hat{\theta})$ reduces $\Psi^I$ and the minimum $F_2$ to zero. In this case, the investor cannot win anything in terms of improved profits by an inappropriate challenge. Finally, the government must be better off by not over-reporting, but revealing the true realization of $\theta$. This condition requires that

$$v(\theta, a^*(\theta)) \geq v(\theta, a_1(\hat{\theta})) + T(\hat{\theta}) - F_1 \tag{8}$$

and is fulfilled due to the specification of $F_1$. Thus, the above mechanism does indeed imply the first best action plan to be implemented.

A crucial feature of the transfer design in ISDS mechanism I is that the transfer is paid for a less ambitious action plan $a_1$. So why is it not possible that the government, after being challenged, will be offered to return to the more ambitious action $a_2$ by *paying*

[10] $F_1 \geq \bar{T}$ also guarantees that the arbitrator will run a budget surplus in case of over-reporting, so the supranational institution will not have to subsidize the arbitration process.

instead of receiving a transfer? Since our design does not rule out that $a_2(\hat{\theta}) = a^*(\hat{\theta})$, this would mean that both parties could return to the optimal plan. However, it is easy to demonstrate that such a mechanism is not feasible: Let $\tau(\hat{\theta})$ denote the transfer from the government to the arbitrator (and ultimately to the investor) for allowing the government to implement action plan $a_2(\hat{\theta})$ instead of $a_1(\hat{\theta})$. In this case, the government will be willing to pay this transfer if

$$v(\theta, a_2(\hat{\theta})) - \tau(\hat{\theta}) > v(\theta, a_1(\hat{\theta})) \Leftrightarrow v(\theta, a_2(\hat{\theta})) - v(\theta, a_1(\hat{\theta})) > \tau(\hat{\theta})$$

holds. However, no $\tau(\hat{\theta})$ exists that will make the government accept this deal if and only if it has over-reported, that is, if and only if $\hat{\theta} > \theta$.[11] If $\tau(\hat{\theta}) = T(\hat{\theta})$ according to (4), the government will reject the deal if if it has over-reported the intervention necessity. Consequently, returning to $a_2$ cannot be part of any feasible mechanism. That still leaves us with a possible design such that the government will refuse to pay a transfer after an appropriate challenge has been made. Therefore, we now consider the following alternative mechanism:

**ISDS mechanism II**

- Stage 1, Stage 2 and Stage 3 as in ISDS mechanism I.
- Stage 4: If $\hat{\theta}$ has been challenged, the government is offered to continue with action plan $a_2(\hat{\theta})$ for which it will *pay* a transfer $\tau(\hat{\theta})$ to the arbitrator. Otherwise action plan $a_1(\hat{\theta})$ is implemented.
  - If the government accepts, action plan $a_2(\hat{\theta})$ is implemented, the government pays $\tau(\hat{\theta})$ to the arbitrator and the investor pays $F_2$ to the arbitrator.
  - If the government rejects, $a_1(\hat{\theta})$ is implemented and the arbitrator pays $F_1$ to the investor .

[11] This follows from the properties of $\phi(x)$ as discussed in footnote 9: the only candidate is $\tau(\hat{\theta}) = \phi(\hat{\theta})$, but due to $v_{xa}(\cdot) > 0$ and $a_2(\hat{\theta}) > a_1(\hat{\theta})$, $\phi(x)$ does not decrease with $x$.

In this setup, accepting the transfer should follow from an inappropriate challenge, and this is the reason why the investor has to pay $F_2$ to the arbitrator if the government is willing to pay this transfer. We find:

**Proposition 2.** *If*

- $\tau(\hat{\theta}) = T(\hat{\theta})$ *according to* (4),
- $F_2 \geq \Psi^I$ , *and*
- $F_1 \geq \Psi^R$ *if* $\Psi^R \geq 0$ *or* $F_1 \geq 0$ *if* $\Psi^R < 0$,

*ISDS mechanism II will imply the optimal action plan* $a^*(\theta)$ *as a subgame-perfect equilibrium.*

The proof is straightforward: (i) As outlined before, the government will accept the alternative plan $a_2(\hat{\theta})$ after an inappropriate challenge only. (ii) An appropriate challenge will be made if $\pi(a_1(\hat{\theta}))+F_1 \geq \pi(a^*(\hat{\theta}))$ which is always fulfilled, and (iii) an inappropriate challenge will not be made if $\pi(a_2(\hat{\theta}))-F_2 \leq \pi(a^*(\hat{\theta}))$ which holds due to the specification of $F_2$. Furthermore, truthful revelation requires $v(\theta, a^*(\theta)) \geq v(\theta, a_1(\hat{\theta})) - F_1$ that is fulfilled due to the specification of $F_1$. Thus, a mechanism in which a transfer is *offered* to the government is also feasible. Note that the requirement for $F_2$ is the same as in ISDS mechanism I, and thus $a_2(\hat{\theta}) = a^*(\hat{\theta})$ will make this requirement also redundant in this setup. We observe that the requirement for $F_1$ is milder because an appropriate challenge will lead to rejection of paying a transfer, whereas the appropriate challenge under ISDS mechanism II implies a transfer from the arbitrator to the government that makes defection more attractive in the first place. In that respect, ISDS mechanism II is easier to implement as it requires a smaller $F_1$.

In any case, the notion of an ISDS compensation in both optimal designs has little to do with ISDS provisions as they are employed in bilateral investment treaties or suggested for multilateral agreements like TTIP and TPP. For example, TPP is designed to indemnify an investor for "unjustified" profit losses. In an optimal design setup, an investor will

not receive any compensation in equilibrium. But even if the government did over-report the intervention necessity, the compensation the investor would receive would be equal to $F_1 - T(\hat{\theta})$ for ISDS mechanism I and $F_1$ for ISDS mechanism II. Both $F_1$ and $T(\hat{\theta})$ are determined by domestic welfare effects only. In particular, any design that wants the government to truthfully report the intervention necessity must rely on domestic welfare effects only. The foreign profit plays a role only for the design of $F_2$ that should keep the investor away from an inappropriate challenge. In that respect, we have seen that specifying $a_2(\hat{\theta}) = a^*(\hat{\theta})$ does not even require any agreement on a payment from the investor to the arbitrator in case of an inappropriate challenge. Therefore, $a_2(\hat{\theta}) = a^*(\hat{\theta})$ makes the design look similar to an ISDS provision that wants the government to compensate the investor, but this compensation is not based on foregone profits. Under ISDS mechanism I, the investor has then the chance to reduce its payment by receiving a transfer, but has to reduce the regulatory burden for the investor from $a^*(\hat{\theta})$ to $a_1(\hat{\theta})$ at the same time. Under ISDS mechanism II, there is also a regulatory relief for the investor, but this comes about because the government is not willing to pay a transfer to return to $a^*(\hat{\theta})$. In any case, the ISDS designs in recent agreements that are based on investor profit losses can never restore efficiency but at best improve the outcome, as shown by Kohler and Stähler (2016).

# 4 An example

The general model has assumed that both the government and the investor can credibly agree on this mechanism including the payments of $F_1$ and $F_2$. Any action plan fulfilling the requirements outlined for Stage 1 of each mechanism will imply certain values for $F_1$ and $F_2$. If the investor and the government are potentially constrained with respect to the payments of $F_1$ and $F_2$, respectively, an important issue for implementation could be that the alternative action plans should not make both $F_1$ and $F_2$ too large in order to guarantee credibility of the agreement. In order to shed more light on these trade-offs, this section considers an example for which domestic welfare is linear-quadratic and foreign

profit is linear in $a$. In particular,

$$v(\theta, a) = \theta a - a^2/2 + \delta - \gamma\theta, \pi(a) = \pi_0 - \beta a, \bar{\pi} = 0, \beta < \underline{\theta}, \delta > \gamma\overline{\theta}, \pi_o > \beta\overline{\theta}, \gamma > 0.$$

A straightforward implication of this specification is that the non-cooperative activity level is given by $\tilde{a}(\theta) = \theta$, as the foreign profit is always positive even without investor protection, and the investment has no value outside the host country, that is, $\bar{\pi} = 0$. The globally optimal activity level is determined by $a^*(\theta) = \theta - \beta$. Note that domestic welfare has also a fixed component $\delta$ that measures the fixed benefit of a foreign investor being active in the domestic country. Additionally, a high $\theta$-realization reduces this benefit directly, but $\delta > \gamma\overline{\theta}$ guarantees that the government has never an interest to see the investor leave the country.

Furthermore, suppose that the government and the investor agree upon a simple reduction method to determine the two alternative action plans:

$$a_i(\hat{\theta}) = a^*(\hat{\theta}) - \alpha_i, a^*(\underline{\theta}) > \alpha_1 > \alpha_2 \geq 0, \overline{\theta} - \underline{\theta} > \beta + \alpha_1. \tag{9}$$

This is a very simple agreement as any investor challenge will imply a reduction in $a^*(\hat{\theta})$ either by $\alpha_2$ or by $\alpha_1$, respectively. Furthermore, (9) assumes that the spread of the $\theta$-realizations is sufficiently large; this assumption will guarantee that $\Psi^R$ will be an interior maximum. The agreement (9) implies that the transfer according to (4) does not depend on $\hat{\theta}$ and is equal to

$$T(\hat{\theta}) = \bar{T} = \frac{(\alpha_1 - \alpha_2)(2\beta + \alpha_1 + \alpha_2)}{2}.$$

This is the transfer that is *offered* to a challenged government for accepting $a_1(\hat{\theta})$ under ISDS mechanism I, and it is the transfer that the government has to *pay* for the action plan $a_2(\hat{\theta})$ under ISDS mechanism II. Note that $\bar{T}$ can be made very small by choosing $\alpha_1$ and $\alpha_2$ to be close to each other. The inappropriate challenge condition requires for both

mechanisms that $F_2 \geq \beta\alpha_2$, so the requirement for $F_2$ becomes smaller with choosing a small $\alpha_2$, and there is no such requirement if $\alpha_2 = 0$. Thus, if the investor is financially constrained and cannot credibly commit to a large $F_2$ payment, no problem arises as long as $\alpha_2 = 0$. The true revelation condition depends on both the true intervention necessity $\theta$ and the announced intervention necessity $\hat{\theta}$. Our specification allows us to scrutinize the optimal defection options of the government. Since transfers do not depend on $\hat{\theta}$ in our example, the government will do best if it chooses $\hat{\theta}$ as to maximize $v(\theta, a_1(\hat{\theta}))$ with respect to $\hat{\theta}$ subject to $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, given that it decides to over-report. Differentiation of $v(\theta, a_1(\hat{\theta}) = \theta(\hat{\theta} - \beta - \alpha_1) - (\hat{\theta} - \beta - \alpha_1)^2/2$ shows that the optimal announced intervention necessity is given by

$$\hat{\theta}^*(\theta) = \min\{\theta + \beta + \alpha_1, \overline{\theta}\}.$$

The optimal defection will always aim at maximizing domestic welfare only, and it can undo the positive effect for the foreign investor and the effect of the alternative action plan by claiming that $\theta$ is larger by $\beta + \alpha_1$ than it actually is. Since the government cannot claim any $\theta$ that is out of range, this over-reporting works only for sufficiently small $\theta$ realizations, but is constrained for large $\theta$ realization by $\hat{\theta} \leq \overline{\theta}$ if $\theta + \beta + \alpha_1 > \overline{\theta}$. We find that

$$v(\theta, a_1(\hat{\theta}^*(\theta))) - v(\theta, a^*(\theta)) = \begin{cases} \beta^2/2 & \text{if } \theta \leq \overline{\theta} - \beta - \alpha_1, \\ \dfrac{\beta^2 - \theta^2 + (\overline{\theta} - \beta - \alpha_1)^2}{2} & \text{if } \theta > \overline{\theta} - \beta - \alpha_1. \end{cases}$$

Note that $v(\theta, a_1(\hat{\theta}^*(\theta))) - v(\theta, a^*(\theta))$ is smaller for $\theta > \overline{\theta} - \beta - \alpha_1$ than for $\theta \leq \overline{\theta} - \beta - \alpha_1$,[12] and thus $\Psi^R = \beta^2/2$; so $\Psi^R$ is independent of $\alpha_1$. The true revelation condition for ISDS mechanism I is given by

[12] $\beta^2/2$ does not change with $\theta$, but differentiating $(\beta^2 - \theta^2 + (\overline{\theta} - \beta - \alpha_1)^2)/2$ with respect to $\theta$ yields $-\theta < 0$. $\overline{\theta} - \beta - \alpha_1 > \underline{\theta}$ due to (9).

$$F_1 \geq \underline{F}^I(\alpha_1, \alpha_2) = \bar{T} + \Psi^R = \frac{(\alpha_1+\beta)^2 - 2\alpha_2\beta - \alpha_2^2}{2},$$
$$\text{where} \qquad \underline{F}^I_{\alpha_1}(\cdot) = \alpha_1 + \beta > 0, \underline{F}^I_{\alpha_2}(\cdot) = -(\alpha_2 + \beta) < 0.$$

Thus, we find that a small $\alpha_1$ will *decrease* and a small $\alpha_2$ will *increase* the necessary payment of the domestic government after an investor challenge. Thus, ISDS mechanism I implies a trade-off between the necessary investor payment $F_2$ and necessary government payment $F_1$ as a small $\alpha_2$ will help the investor in terms of a moderate $F_2$ requirement. We may expect that financial constraints may be more likely to be binding for the investor than for the government, and in this case, $\alpha_2 = 0$ will do the job. The necessary $F_1$ can be reduced by selecting an $\alpha_1$ not too much larger than $\alpha_2$.

The true revelation condition for ISDS mechanism II is less demanding and given by

$$F_1 \geq \underline{F}^{II} = \Psi^R = \beta^2/2.$$

Therefore, mechanism II could be preferred if the domestic government is also facing financial constraints that make an upfront payment of $\underline{F}^I(\alpha_1, \alpha_2)$ difficult. In any case, our example has shown that very simple rules can be applied if domestic welfare can be approximated by a linear-quadratic function and maximized profits can be approximated by a linear function.

# 5 Concluding remarks

This paper has developed two different types of an optimal ISDS mechanism. This mechanism has little in common with the recent designs as found in bilateral and multilateral investment treaties. First, any ISDS compensation should be based on the host country's welfare effects, and not on any foregone profit. Second, an efficient investment protection provision must involve three parties and therefore needs to be managed credibly by supranational institutions. Recent designs fail on these accounts, and can, at best, only

improve outcomes.

The paper does in no way indicate that the suggested ISDS mechanism is politically feasible. It requires a thorough cost-benefit analysis within host countries, commitment to a transfer scheme and action plans, and it relies on arbitration. However, recent designs that employ ISDS tribunals also require commitment to these ISDS rules, so any investor protection will be infeasible without a minimum degree of commitment. Furthermore, arbitration may make any investor protection provision more feasible than a scheme that will have to rely on the judgement of tribunals that will not follow well-defined rules. A further caveat for the optimal mechanism is that the investor must be able to learn the intervention necessity. But if she cannot (and therefore the proposed mechanism is not feasible), it is hard to see how a tribunal can do a better job. It also seems to be a fair assumption that investors know the impact of their activities not only on their profits, but also on other agents.

The paper has borrowed parts of the mechanism design from the implementation literature. However, as discussed by Maskin (2002), these subgame-perfect implementation mechanisms are hardly ever used to tackle the holdup problem between private agents, so a market in which arbitrators offer their services to solve the holdup problem efficiently does obviously not exist. He concludes that there must be other frictions that play a role for the non-existence of these markets. In our context, however, supranational institutions that can take over the role of an arbitrator already exist. For example, the International Centre for Settlement of Investment Disputes (ICSID), a part of the World Bank Group, is such an institution that has already experience in investment dispute resolution. Moreover, the WTO is already an expert in arbitration through its Dispute Settlement Body (DSB) that deals with trade disputes. It thus seems that supranational institutions must be given an active role to make investor protection successful and acceptable.

# References

[1] Aghion, P., Fehr, E., Holden, R., Wilkening, T. (2015), The Role of Bounded Rationality and Imperfect Information in Subgame Perfect Implementation – An Empirical

Investigation, IZA Discussion Paper No. 8971.

[2] Aisbett, E., Karp, L., McAusland, C. (2010a), Compensations for Indirect Expropriation in International Investment Agreements: Implications for National Treatment ad Rights to Invest, *Journal of Globalization and Development*, 1 (2), Article 6.

[3] Aisbett, E., Karp, L., McAusland, C. (2010b), Police Powers, Regulatory Takings and the Efficient Compensation of Domestic and Foreign Investors, *Economic Record*, 86, 367-383.

[4] Antràs, P. (2003), Firms, Contracts, and Trade Structure. *Quarterly Journal of Economics*, 118, 1375-1418.

[5] Antràs, P., Helpman, E. (2004), Global Sourcing. *Journal of Political Economy*, 112, 552-80.

[6] Antràs, P. , Chor, D. (2013), Organizing the Global Value Chain, *Econometrica*, 81, 2127-2204.

[7] Bagwell, K. Staiger, R.W. (2005), Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO, *Journal of International Economics*, 67, 268-294.

[8] Bagwell, K., Bown, C.P., Staiger, R.W. (2016), Is the WTO passé?, *Journal of Economic Literature*, forthcoming.

[9] Berger, A., Busse, M., Nunnenkamp, P., Roy, M. (2011), More stringent BITs, less ambiguous effects on FDI? Not a bit! *Economics Letters*, 112, 270-272.

[10] Blume, L., Rubinfeld, D., Shapiro, P. (1984), The taking of land: When should compensations be paid? *Quarterly Journal of Economics*, 99, 71-92.

[11] Egger, P., Merlo, V. (2012), BITs Bite: An Anatomy of the Impact of Bilateral Investment Treaties on Multinational Firms, *Scandinavian Journal of Economics*, 114, 1240-1266.

[12] Hermalin, B. (1995), An economic analysis of takings, *Journal of Law, Economics and Organization*, 11, 64-86.

[13] Kohler, W., Stähler, F. (2016), The Economics of Investor Protection: ISDS versus National Treatment, CESifo Working Paper No. 5766.

[14] Maskin, E. (2002), On Indescribable Contingencies and Incomplete Contracts. *European Economic Review*, 46, 725-733.

[15] Maskin, E., Tirole, J. (1999), Unforeseen Contingencies and Incomplete Contracts, *Review of Economic Studies*, 66, 83-114.

[16] Moore, J., Repullo, R. (1988), Subgame perfect implementation, *Econometrica*, 56, 1191-1220

[17] Navaretti, G.B., Venables A.J. (2006), Multinational Firms in the World Economy, Princeton University Press.

[18] Noval, E. (2001), The taking of land: market value compensation should be paid, *Journal of Public Economics*, 82, 431-443.

[19] OECD (2012), Investor-State Dispute Settlement, Public Consultation: 16 May - 9 July 2012, Paris.

[20] TPP draft (2016), Text of the Trans-Pacific Partnership, New Zealand Ministry of Foreign Affairs and Trade, http://tpp.mfat.govt.nz/text, accessed January 26, 2016.